

# 새싹 플레이리스트



곡 장르와 태그를 기반으로  
플레이리스트를 추천하는 시스템 개발

# 프로젝트 참여 인원

아래 총 4명의 새싹 교육생이 참여함



윤혜영

데이터 전처리 임무



조경희

콘텐츠 기반 필터링 임무



장윤식

협업 필터링 (baseline 확인) 임무



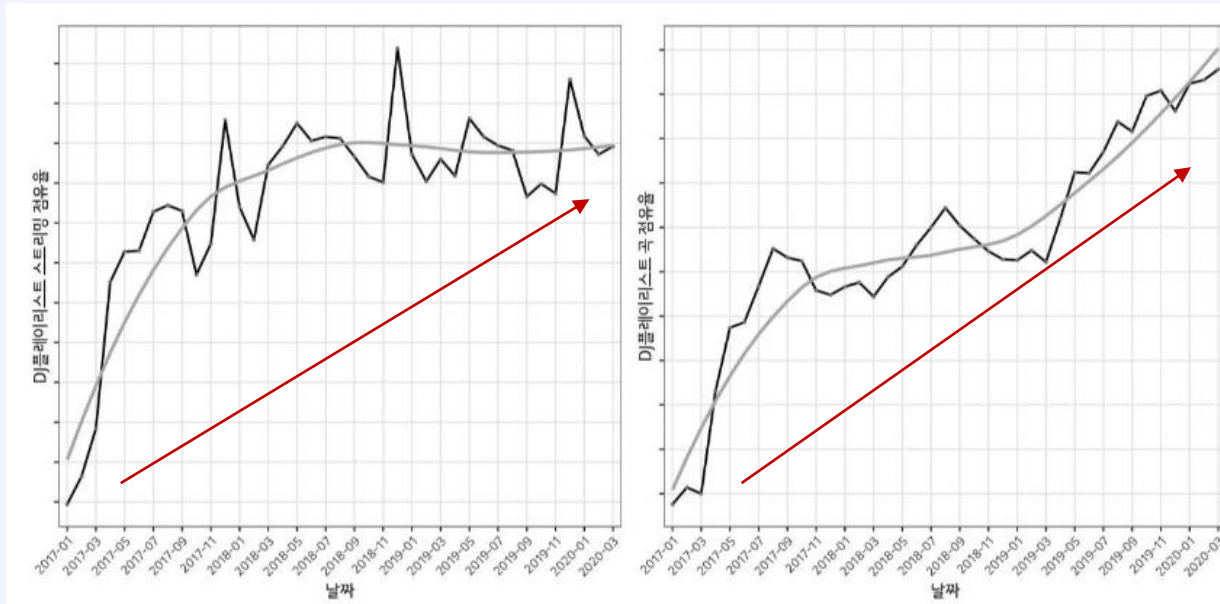
김 빈

자연어 처리 임무

# 프로젝트 배경

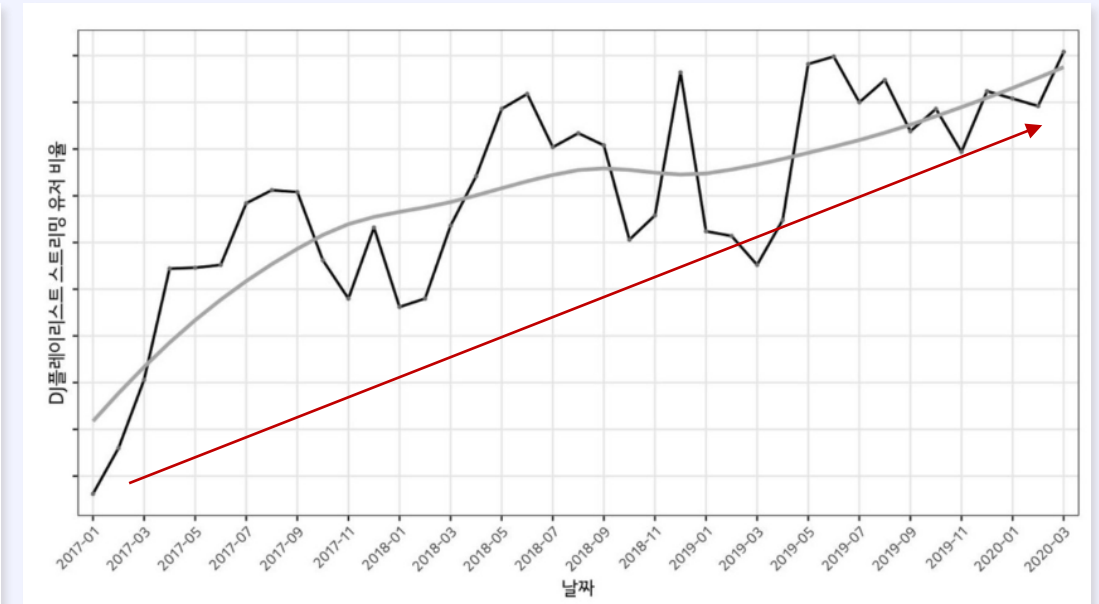
‘내가 선호하는 노래’를 일일이 찾는 수고로움에서 벗어나 취향에 맞는 노래를 편하게 듣고 싶어하는 MZ세대.  
이들은 유튜브 플레이리스트 또는 음원 사이트의 특정 플레이리스트를 선호하는 경향을 보임

(자료 출처: 멜론 플레이리스트 데이터 - 카카오 제공)



일 평균 DJ 플레이리스트 스트리밍 점유율  
(2017/01 ~ 2020/03 월 단위)

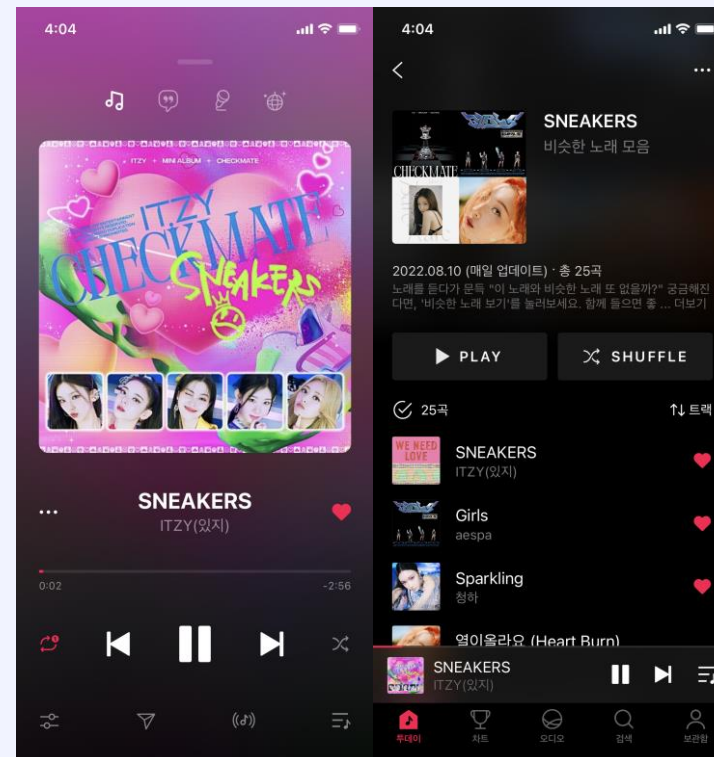
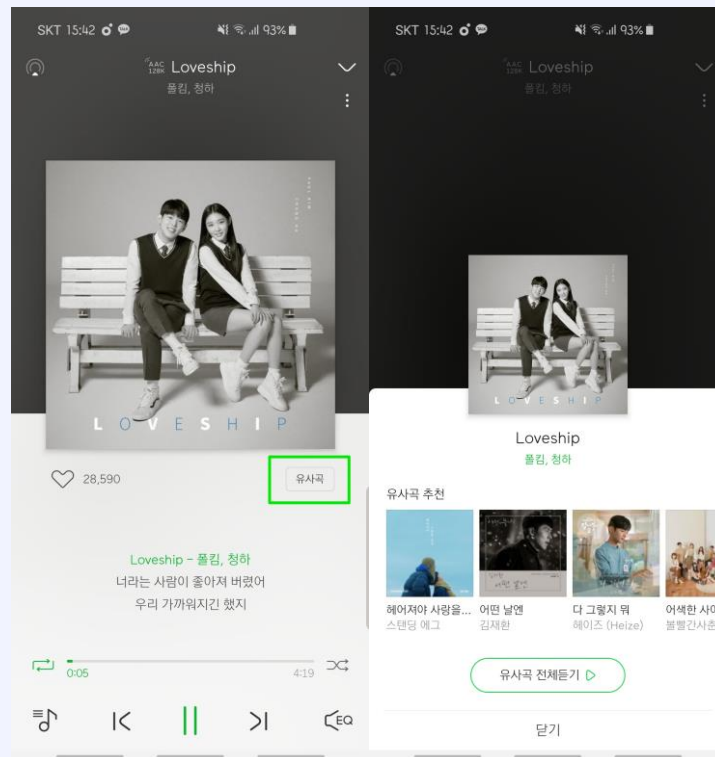
일 평균 DJ 플레이리스트 곡 점유율  
(2017/01 ~ 2020/03 월 단위)



일 평균 DJ 플레이리스트 스트리밍 유저 비율  
(2017/01 ~ 2020/03 월 단위)

# 프로젝트 배경

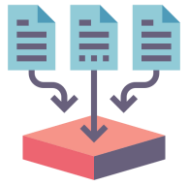
현재의 음원 사이트/어플은 여전히 곡 기반의 추천 시스템을 유지하고 있는 상황으로,  
세대의 특성에 맞는 새로운 형태의 추천 시스템(플레이리스트 기반 추천 시스템)을 개발하고자 함



# 프로젝트 진행 절차

아래 절차는 Rutgers University에서 개발한 하이브리드 추천 시스템 개념도를 참고하여 작성했습니다.

1



## 데이터 수집

- 멜론 곡/아티스트 정보
- 멜론 곡 태그 정보

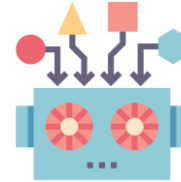
2



## 데이터 속성 추출

곡/아티스트별 속성 추출

3



## 모델 학습

Content-based Filtering  
- K-Mean Clustering  
- feature별 유사도 분석  
(코사인 유사도, 태그간 유사도)

4



## 결과 도출

태그/장르에 맞는 음악 추천  
(태그로 분류되지 않은 단어를  
입력하더라도 추천 음악 도출)

## 프로젝트 진행 절차 - 1) 데이터 수집

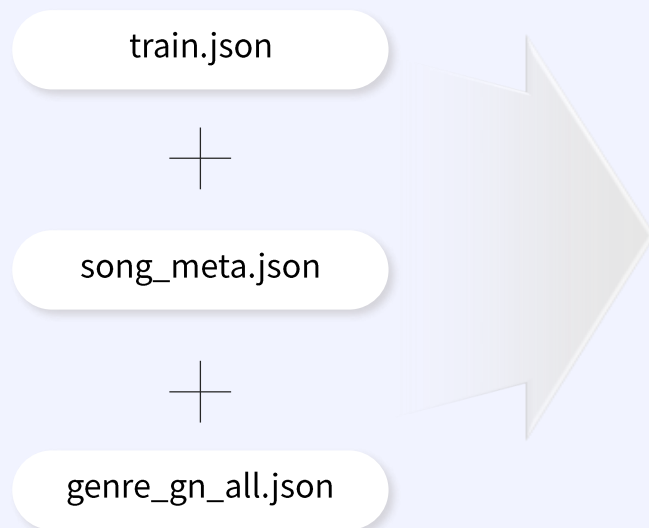
카카오 아레나에서 다운받을 수 있는 멜론 데이터를 이용할 계획 - train.json, song\_meta.json, genre\_gn\_all.json



## 프로젝트 진행 절차 - 2) 데이터 속성 추출



앞서 수집한 3가지 데이터를 결합 및 전처리하여 하나의 데이터셋을 구축한 후, 필요한 데이터 속성을 추출할 계획



	tags	id	plylst_title	songs	like_cnt	updt_date	song_gn_gnr_basket	song_gn_dtl_gnr_basket
0	[락]	61281	여행같은 음악	[Hey Little Girl, Octagon, The Road, Honeymoon...	71	2013-12-19 18:36:19.000	[GN1400, GN1000, GN1100, GN1300, GN1900, GN0900]	[GN1001, GN1003, GN1901, GN1301, GN1101, GN140...
1	[추억, 회상]	10532	요즘 너 말야	[한사람을 위한 마음, Audition (Time2Rock), 기 다리다, 좀 더 룰...	1	2014-12-02 16:19:42.000	[GN0100, GN1000, GN1800, GN0800, GN1700, GN260...	[GN0104, GN1502, GN1706, GN0103, GN2602, GN180...
2	[카페, 잔잔한]	76951	편하게, 잔잔하게 들을 수 있는 곡..	[도시의 밤, I'm Alright, 너를 좋아하니까, 247 (Feat. AMJ)...	17	2017-08-28 07:09:34.000	[GN0100, GN0300, GN0800, GN1700, GN2600, GN040...	[GN0403, GN0401, GN0303, GN0301, GN0606, GN040...
3	[연말, 눈오는날, 캐럴, 분위기, 따뜻한, 크리스마스캐럴, 겨울노래, 크리스마스...	147456	크리스마스 분위기에 흥분 취하고 싶을때	[Into the Unknown (From &#34;Frozen 2&#34;/Sou...	33	2019-12-05 15:15:18.000	[GN0100, GN2200, GN1800, GN1300, GN0800, GN170...	[GN1706, GN0103, GN0506, GN0908, GN1506, GN180...
4	[댄스]	27616	추억의 노래	[눈물에 얼굴을 묻는다, 타인, 학원별곡 (學園別曲), 날 떠나지마, No.1, D...	9	2011-10-25 13:54:56.000	[GN0100, GN0300, GN2500, GN0200, GN0400, GN0600]	[GN0302, GN0403, GN0601, GN0103, GN2506, GN250...



## 프로젝트 진행 절차 - 3) 모델 학습

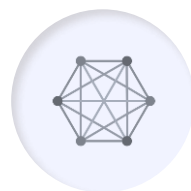
콘텐츠 기반 필터링의 방식을 이용해, 대상 자체의 특성을 바탕으로 추천 플레이리스트를 도출하는 모델 구축 예정



### 벡터 표현

곡을 표현할 수 있는 단어를  
벡터로 변환하는  
Word2Vec을 사용

- 사전 학습된 Word2Vec을  
fine-tuning하여 사용



### 유사도 분석

벡터의 유사도 측정을 위해  
코사인 유사도를 사용

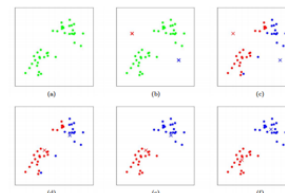
$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$


$$= \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$



### 군집화

유사한 데이터의 군집화를 위해  
K-mean clustering 사용



 곡, 태그를 바탕으로  
플레이리스트 추천

태그 간 유사도를 계산하여  
유사한 태그를 도출할 수 있으며,  
태그 간 연산(더하기, 빼기)을 통해  
관련성 있는 태그를 확인할 수 있음

태그 간 유사도	태그 Composability - 더하기?
세련된 느낌있는 편협성 트렌디 감각적인 유니크 느낌 매력적 groove	미련 연애 헤어짐 보고싶다 이별후 아름 고백 썸사랑 후회중
쿨 + 후회	

곡과 태그를 입력 받아  
추천 플레이리스트를 도출하는 모델 구축

(참고: <https://wikidocs.net/22660>)



## 프로젝트 진행 절차 - 4) 결과 도출

입력한 태그 및 곡에 맞는 플레이리스트를 추천하며,  
궁극적으로는 사전에 태그로 정의하지 않은 단어를 입력하더라도 추천 결과를 도출하는 모델을 목표로 하고 있음



팝송 **Playlist 2022** ❤️ 나는 오늘 기분이 좋다 | 가볍고 시끄럽지 않은 느낌 🎵 🎵 **pop r&b mix** 🍃  
조회수 3.3천회 · 12시간 전

 오늘의 Chill Playlist

팝송 Playlist 2022 나는 오늘 기분이 좋다 | 가볍고 시끄럽지 않은 느낌 🎵 🎵 pop r&b mix TRACKLIST ...

새 동영상



**Playlist 2022** ❤️ 기분이 좋아지는 진짜 좋은 상쾌한 팝송 - 산뜻한노래, 행복 🎵 🎵 **pop music** ✨

343명 시청 중

 오늘의 Chill Playlist

Playlist 2022 기분이 좋아지는 진짜 좋은 상쾌한 팝송 - 산뜻한노래, 행복 🎵 🎵 pop music ✨ TRACKLIST ...

🔴 실시간 새 동영상